Institut
de recherche
et d'information
sur la rémunération

METHODOLOGIE STATISTIQUE DE L'ENQUETE DU CRSMT

Etude de faisabilité

Document à l'usage exclusif des parties syndicales et patronales des secteurs public et parapublic du Québec

AVANT-PROPOS

En novembre 1985, l'IRIR faisait l'inventaire des sources sur la rémunération susceptibles d'être utilisées aux fins d'une comparaison entre les salariés des secteurs public et parapublic et les autres salariés québécois. L'enquête du CRSMT avait alors été jugée une piste prometteuse. Cependant l'Institut s'était réservé le droit d'analyser en profondeur cette méthode afin de pouvoir juger de sa pertinence et de sa fiabilité.

Ce rapport vise donc à analyser la méthodologie de l'étude du CRSMT dans une perspective de rémunération. Il présente les volets principaux de cette méthode en faisant l'analyse des avantages et inconvénients de chacun. Il fait également état de certaines vérifications qui pourraient être entreprises par le CRSMT.

Ce rapport a été préparé par Christopher Thom, analyste en rémunération à l'IRIR sous la direction de Kevin O'Farrell.

Luc Bessette
Vice-président et
directeur de la recherche

Juin 1986

TABLE DES MATIERES

	Page
Méthodologie.....	1
Vue d'ensemble de la méthodologie utilisée.....	2
L'analyse de la méthode du CRSMT.....	4
-Définition de l'univers.....	4
-Choix des strates.....	7
-Echantillonnage.....	9
-Pondération des données.....	12
-Vérification de la fiabilité des données.....	17
Conclusion.....	25
Annexes.....	28

METHODOLOGIE

Afin de préparer ce rapport, les démarches suivantes ont été entreprises:

1. Consultation de sources écrites (ouvrages théoriques, académiques et techniques);
2. Discussions avec les personnes ressources au CRSMT, à Statistique Canada, au Bureau of Labour Statistics à Washington, et avec les spécialistes externes et les analystes de l'IRIR.

VUE D'ENSEMBLE DE LA METHODOLOGIE UTILISEE

Le CRSMT a comme objectif de recueillir des données sur des emplois-repères dans des établissements ayant plus de 200 employés au Québec. Afin d'accomplir ce mandat tout en minimisant les coûts, le CRSMT applique une méthodologie qui résulte dans l'échantillonnage aléatoire des établissements ayant de 200 à 499 employés et ceux du secteur manufacturier ayant 500-999 employés. Il fait ensuite le recensement de toutes les autres catégories d'établissements.

Dans la plupart des sondages, l'unité échantillonnée possède la donnée recherchée. Dans le cas d'un échantillon pour connaître le salaire moyen dans une entreprise, cette information sera disponible en autant que l'entreprise a des employés. Par contre, lorsque l'on cherche le salaire moyen pour un type d'emploi particulier, la donnée ne sera disponible que si cet emploi existe au sein de l'entreprise échantillonnée.

La méthode utilisée par le CRSMT, fait l'estimation de la rémunération d'un ensemble d'établissements basée sur des observations recueillies dans un nombre restreint d'établissements. Ces estimations, même en répondant à des critères sévères de précision et d'exactitude, produisent toutefois une marge d'erreurs qui peut être problématique dans le cadre d'une comparaison de la rémunération.

Prenons l'exemple où des données obtenues et pondérées produisent un taux moyen de salaire de 30 000$. En appliquant les critères du CRSMT et pourvu qu'il n'y ait pas d'erreurs dans le processus d'échantillonnage, d'appariement, de traitement de données, etc., on aurait 95% de chances que le taux moyen de salaire se situe entre 27 000$ et 33 000$, c'est-à-dire à 30 000$ plus ou moins 10% (2 fois le coefficient de variation de 5%) et avec un biais relatif de 300.00$ (1).

Etant donné qu'une enquête salariale comporte toujours des possibilités d'erreurs humaines dans le choix des emplois-repères, l'appariement des emplois et la fiabilité des informations reçues des établissements, et que la méthode par échantillonnage aléatoire possède une marge d'erreurs inhérente, la façon dont la méthodologie est utilisée devient extrêmement importante.

---

(1) Le CRSMT établit son niveau de confiance à 95%, ses critères d'acceptation et le coefficient de variation à 5%, et le biais relatif à 1%.

Une enquête par échantillonnage aléatoire comporte toujours une marge d'erreurs en raison des estimations inhérentes. Un niveau de confiance de la représentativité de l'échantillon par rapport à l'univers est établi afin de déterminer la grandeur appropriée de l'univers. Des mesures d'exactitude et de précision des données (le coefficient de variation et le biais relatif) sont calculées après que les données observées aient été recueillies et pondérées, en estimant leur écart par rapport à l'univers et à leur biais.
Voir annexe 1.

L'ANALYSE DE LA METHODE DU CRSMT

Cette analyse de la méthodologie de l'enquête du CRSMT se divise en cinq étapes: la définition de l'univers, le choix des strates (stratifications), l'échantillonnage, la pondération des données et finalement le calcul des mesures de fiabilité et de précision. Les trois premières étapes expliquent la détermination des établissements ou unités répondantes dans lesquelles l'enquête sera conduite. Les deux dernières étapes décrivent les modes de calcul et de vérification des données recueillies par l'enquête.

Choix des établissements
- -Définition de l'univers
- -Choix des strates
- -Echantillonnage

Production des résultats
- -Pondération des données
- -Vérification de la fiabilité des résultats

1. Définition de l'univers:

1.1 Principes

L'univers d'une enquête en rémunération peut se composer soit d'entreprises ou d'établissements. Le choix entre les entreprises et les établissements comme unité de base est fonction de certains critères tels: l'exclusivité des unités ou le besoin de précision ou de représentativité sur le plan sectoriel ou régional. L'univers doit être approprié, complet, mis à jour, et ce en fonction des objectifs de l'enquête.

1.2 Analyse

L'univers du CRSMT est composé de 856 établissements, et inclut tous les établissements ayant plus de 200 employés, sauf dans les cas des municipalités où le critère minimal exige plus de 25,000 habitants. Par exemple, l'enquête inclut six municipalités qui ont plus de 25,000 habitants mais qui emploient moins de 200 employés.

Le CRSMT enquête aussi les entreprises multi-tailles, qui comprennent au moins un établissement de 200 employés et plus, ainsi que d'autres établissements qui peuvent avoir également au moins 200 employés mais pas nécessairement. Ces établissements sont regroupés dans cette catégorie parce qu'ils ne peuvent rapporter les données pour un seul établissement avec leur système de gestion interne. Par cette approche, même si une entreprise a plus de 200 employés, seuls les établissements ayant 200 employés et plus sont couverts par le CRSMT. Par conséquent, un certain nombre d'établissements dans des secteurs comme les banques ou les chaînes d'alimentation ayant plusieurs unités ne sont pas comptés dans l'univers.

Le CRSMT établit son univers en début d'année à partir de listes fournies par la CSST, Statistique Canada et le CRIQ. Les établissements qui sont considérés problématiques pour différentes raisons (ceux qui ne sont plus dans l'univers à cause d'une réduction d'effectifs ou d'une faillite, etc.) sont aussi contactés en début d'année.

Cette vérification permet de corriger et de mieux préciser l'univers avant que l'enquête commence au mois d'avril. Cependant, des changements continus dans l'économie et dans les entreprises font que cette vérification n'est ni complète ni à jour. Par exemple, souvent les enquêteurs du CRSMT ne constatent des changements à l'univers que lors de l'enquête elle-même. Ceci peut engendrer un problème au niveau de la pondération et aussi au niveau des résultats, dépendant de la façon que l'on traite ces inexactitudes. Ce point sera discuté ultérieurement.

1.3 Avantages et inconvénients

Avantages

1.3.1 Le CRSMT couvre plus d'établissements que toute autre enquête en rémunération au Québec en incluant des établissements de 200-499 employés et des administrations publiques et para-publiques.

1.3.2 L'utilisation de l'établissement comme unité de base est appropriée car elle permet une certaine précision et une représentativité sur le plan sectoriel et régional, et donne plus de flexibilité dans la présentation des résultats.

Inconvénients

1.3.4 L'utilisation de l'établissement comme unité de base exclut certaines entreprises qui n'emploient pas 200 employés dans au moins un de leurs établissements.

2. Choix des strates

2.1 Principes

Dans l'échantillonnage aléatoire, la stratification sert à réduire les variances en regroupant les unités homogènes et à identifier le nombre optimal d'établissements à enquêter pour atteindre un niveau de précision acceptable.

2.2 Analyse

Le CRSMT a choisi trois catégories de stratification: le secteur d'activité économique, la taille d'établissement (nombre d'effectifs) et la région géographique. Le CRSMT a ventilé la stratification du secteur économique selon les catégories couramment utilisées par le Bureau de la Statistique du Québec et Statistique Canada.

La ventilation par taille d'établissement se fait selon les six catégories d'effectifs suivantes: 200-499, 500-999, 1000-1499, 1500-2499, 2500-5000, 5000+, et multi-tailles dans le cas du volet "salaires". Le CRSMT a gardé les strates de taille afin de pondérer les non-réponses d'une façon plus exacte. Il est à noter aussi que ces ventilations sont réduites à quatre dans la présentation du volet "rémunération globale" à cause d'un taux de réponse plus faible dans l'enquête sur les avantages sociaux.

La ventilation par région administrative se fait selon quatre catégories: Montréal, Québec, autres régions et multi-régions. Cette ventilation régionale a été abandonnée dans le calcul des avantages sociaux à cause du faible taux de réponse. Il est à noter que les résultats sont publiés dans Le Marché du travail selon le secteur économique, la taille et la syndicalisation, mais pas selon la région géographique.

La stratification prend son importance du fait qu'elle détermine la méthode d'échantillonnage à utiliser. Tel que mentionné, tous les établissements du secteur privé ayant 200-499 employés, et du secteur manufacturier ayant 500-999 employés, sont soumis à l'échantillonnage aléatoire. Les autres strates sont recensées à cause du nombre trop restreint d'établissements au sein de chacun.

Echantillonnage	No. d'établissements
Aléatoire	686
Recensé	170
---	---
Univers	856

Afin de s'assurer d'une représentativité de son univers sans faire un recensement complet et en même temps de minimiser les coûts de l'enquête, le CRSMT utilise un échantillonnage aléatoire dans les stratifications ayant le plus d'établissements.

2.3 Avantages/Inconvénients

Avantages

1. La stratification retenue offre un grand nombre de ventilations pour l'analyse des données sur la rémunération.
2. La stratification est complète et regroupe les mêmes divisions que la plupart des enquêtes de cette envergure (ex. Statistique Canada).

3. Echantillonnage

3.1 Principes

Deux objectifs d'une enquête par échantillonnage aléatoire sont d'atteindre la précision désirée et de minimiser le nombre d'établissements à enquêter pour des raisons de coûts.

La méthode optimale de Neymann utilisée par le CRSMT est reconnue comme étant fiable (voir Sheaffer pp. 59-122, et Cochran pp. 72-169). Cette méthode détermine le nombre optimal d'établissements à enquêter basé sur la variabilité des salaires dans une ou plusieurs strates. Afin de calculer le nombre optimal d'établissements, il faut connaître la variance des salaires de chacun des secteurs principaux. La formule de Neymann détermine le nombre d'établissements qui devraient être enquêtés par secteur principal. Ensuite le nombre d'établissements de chaque secteur principal peut être alloué aux sous-stratifications de façon proportionnelle.

Après la sélection des établissements, on constate, dans le cas où toutes les hypothèses et calculs sont fiables, que les établissements sélectionnés représentent tous les établissements de l'univers au niveau de confiance établi (95%).

Ainsi, pour établir le nombre optimal d'établissements à enquêter, il faut calculer les variances des salaires des emplois-repères des stratifications visées d'une façon très précise afin que tous les emplois-repères soient adéquatement échantillonnés. Les variances sont essentielles pour assurer la représentativité et la fiabilité des résultats.

3.2 L'analyse

Le CRSMT se sert de la méthode optimale de Neymann dans son échantillonnage stratifié dans le cas des résultats globaux présentés dans l'encart de la revue le Marché du Travail. Dans la ventilation public/privé, il utilise l'échantillonnage stratifié des domaines pour tenir compte des établissements publics et privés à l'intérieur des mêmes strates (voir Cochran, pp. 144-147).

L'échantillonnage se fait sur la base optimale selon le secteur économique sauf dans le cas des industries manufacturières où les deux tailles (200-499, 500-999) sont traitées séparément. L'allocation des établissements par région est faite ensuite de façon proportionnelle. Pour s'assurer que tous les emplois-repères sont adéquatement représentés, le CRSMT fait l'addition de toutes les variances d'un emploi-repère par secteur économique, et ensuite il calcule la moyenne de la somme des variances de tous les emplois-repères par secteur économique.

Le CRSMT ajuste son univers quant à l'inclusion ou l'exclusion d'établissements par rapport aux critères tels que le nombre d'effectifs. La plupart des établissements enquêtés l'année précédente qui se conforment encore aux paramètres de l'univers sont enquêtés à nouveau de façon automatique. Un biais se crée ainsi car tous les établissements ne sont pas sélectionnés au hasard à toutes les années. Le Bureau of Labour Statistics à Washington refait son échantillon au complet à tous les trois ans afin de s'assurer d'un échantillonnage réellement aléatoire. A l'intérieur de cette période, des tendances peuvent être surveillées.

Deuxièmement, les variances utilisées pour calculer le nombre minimal et optimal des établissements enquêtés sont basées en partie sur l'estimation des variances qui a été faite lors de la constitution du premier échantillonnage en 1982. Même si les variances sont recalculées à chaque année pour tenir compte de ceux qui entrent et sortent de l'univers, elles sont faites en grande partie à partir des données de l'année précédente. Puisque tous les établissements ne sont pas sélectionnés au hasard à chaque année, les premières estimations pourraient continuer à avoir un certain impact aujourd'hui.

3.3 Avantages/Inconvénients

Avantages

1. L'enquête utilise la méthode optimale et proportionnelle de Neymann qui est reconnue et fiable.
2. Les tendances salariales peuvent être analysées à partir des établissements qui sont enquêtés d'une année à l'autre.

Inconvénients

1. Il n'y a pas de ré-échantillonnage complet de l'univers. Un tel re-échantillonnage pourrait maintenir l'intégrité à long terme de la technique aléatoire en plus d'aider à confirmer la validité de la prémisse que les taux de rémunération des emplois dans les établissements enquêtés représentent les taux de rémunération des emplois dans les établissements non-enquêtés.

4. Pondération des données

4.1 Principes

Les données des établissements enquêtés sont pondérées afin de représenter tous les établissements de leurs strates, en vue d'établir l'effet d'un recensement. Prenons l'exemple où une strate est composée de huit établissements et où trois établissements ont été échantillonnés. Les données des trois établissements répondants sont pondérées par 8/3 afin de produire le même effet que si tous les huit établissements étaient enquêtés.

Pour que cette pondération soit exacte, il faut que l'échantillonnage soit complet, précis et à jour, que les appariements lors des enquêtes soient bien faits, et que les données des emplois-repères enquêtés représentent les données des emplois-repères non-enquêtés.

4.2 L'analyse

La pondération des données observées peut produire des résultats différents selon que l'on procède par recensement ou échantillonnage aléatoire. Voici un exemple.

Recensement: dans la stratification fictive qui suit, parmi les 8 établissements dans l'univers et dans l'échantillon recensé, seulement 3 ont répondu à l'étude du CRSMT.

Univers:	8
Echantillonnage:	8
Réponses	3

L'univers est divisé par la réponse afin d'établir la pondération:

$$\text{Pondération} = \frac{\text{Univers}}{\text{Réponse}}$$

$$= 8/3$$

Les données de chacun des établissements répondants sont ainsi pondérées par 8/3 afin de créer l'effet que les huit établissements ont répondu à l'étude.

Echantillonnage aléatoire: dans cette deuxième stratification fictive, la taille de l'univers et les réponses sont les mêmes que dans l'exemple du recensement. Cependant, l'échantillonnage est composé de quatre établissements plutôt que huit.

Univers: 8

Echantillonnage: 4

Réponses 3

Ici il faut pondérer les réponses par rapport à l'univers et par rapport à l'échantillon. Le premier est pondéré parce que l'échantillon est calculé pour représenter l'univers. Le deuxième est pondéré parce que les données et les caractéristiques des réponses sont supposées représenter les données et les caractéristiques des établissements à l'intérieur de l'échantillon.

Dans le cas des échantillons aléatoires, le quotient Univers/Réponse est multiplié par le quotient Réponse/Echantillonnage. Par exemple:

$$\text{Pondération} = \frac{\text{Univers}}{\text{Réponse}} \times \frac{\text{Réponse}}{\text{Echantillon}}$$

$$= 8/3 \times 3/4$$

$$= 2$$

Puisque seulement trois quarts de l'échantillon existent en réalité, soit à cause d'un établissement qui fait faillite ou qui a réduit ou augmenté son personnel de façon à changer de strate, et que l'échantillon de 4 établissements est supposé représenter l'univers, le CRSMT émet l'hypothèse que l'univers peut être modifié dans la même proportion que l'échantillon: de 8 à 6 dans l'univers, comme de 4 à 3 dans l'échantillon. Ainsi une réponse de 3 dans un univers de 6 donnerait une pondération de 6/3, et non pas de 8/3.

L'allocation des établissements dans les échantillons stratifiés par le CRSMT se fait sur la variance des salaires moyens. Le fait d'ajuster la taille de l'univers en fonction de la taille de l'échantillon est discutable. Dans l'exemple aléatoire ci-haut, supposons que la réduction du nombre d'établissements est dûe à un décès (faillite). Il n'est pas certain qu'il y ait eu un décès parmi les 4 établissements qui n'ont pas été sélectionnés. Dans le cas de non-réponse, il n'est pas certain que les salaires des 3 établissements qui restent puissent représenter de façon précise les salaires de l'établissement manquant.

Ainsi, trop ou pas assez de poids pourra être accordé aux données observées d'une stratification, ce qui entraînera des résultats biaisés. Prenons le cas où il y a beaucoup de décès dans les échantillons d'une stratification aléatoire, mais où il n'y a pas de décès parmi ceux de cette stratification qui ne sont pas échantillonnés. La méthode utilisée présentement par le CRSMT réduirait le taille de l'univers de la stratification et accorderait moins de poids aux établissements enquêtés. Ceci aurait l'effet de biaiser les résultats par rapport à celles des autres stratifications ainsi que les résultats au niveau provincial.

4.3 Avantages/Inconvénients

Avantages

1. La pondération crée l'effet d'un recensement complet.

Inconvénients

1. Un biais dans les résultats peut être créé quand les données sont pondérées, en raison de l'absence de preuve quant à la prémisse que les établissements enquêtés représentent les établissements non-enquêtés, et en raison du fait que le CRSMT n'ajuste pas son univers lors d'une non-réponse. Ce biais peut être plus grand dans le cas "aléatoire" parce que les données sont pondérées deux fois: la première pour rétablir l'échantillon et la deuxième pour pondérer l'échantillon au niveau du recensement.

Dans la mesure du possible, le CRSMT devrait enquêter d'autres établissements lors de refus. Il devrait aussi ajuster le grandeur de l'univers dans ces calculs pour tenir compte des décès et changements de taille.

5. Vérification de la fiabilité des données

5.1 Principes

La précision et l'exactitude des données obtenues et pondérées sont estimées par le CRSMT en utilisant le coefficient de variation et le biais relatif.

Coefficient de variation (CV): le coefficient de variation indique l'erreur de l'échantillonnage. Le CV est basé sur l'hypothèse que plus petite est la variabilité des données recueillies par échantillon aléatoire plus près sera la moyenne des estimations de la moyenne de l'univers (voir Cochran, pp. 155-156).

En faisant le calcul du CV, on peut déterminer si les estimations des taux moyens de salaires des établissements non-enquêtés sont précises. Selon Cochran, un CV de moins de 10% signifie un niveau de confiance acceptable en ce qui concerne la précision des résultats.

Biais relatif (BR): Le biais relatif est un indicateur du biais systématique dans certaines estimations. Un biais ne peut pas être calculé dans le cas où il provient de mauvais appariements ou d'un manque de données. Il peut être calculé cependant quand l'estimateur utilisé est par définition biaisé. Un exemple d'un tel estimateur est celui du taux moyen de salaire tel qu'utilisé dans l'étude du CRSMT. Le genre de formule utilisée par le CRSMT, qui est appelée un estimateur quotient combiné, a comme dénominateur une estimation:

$$\frac{\text{masse salariale}}{\text{nombre d'effectifs dans les emplois-repères}}$$

C'est cette estimation du nombre d'effectifs dans les établissements non-enquêtés qui produit la possibilité d'un biais relatif systématique, lequel peut être calculé par la méthode "Série de Taylor" (voir Cochran).

5.2 L'analyse

L'enquête du CRSMT se sert de ces deux mesures statistiques comme indicateurs de rejet pour les emplois-repères.

Aucun emploi-repère n'est rejeté à cause du CV, mais cinq emplois-repères sont rejetés à cause des BR trop élevés:

Rejets des emplois-repères

Emploi-repère	Variable	Raison
Gardien	Sal Moyen	BR = D
" "	Rém Glob Min	BR = C
" "	Rém Glob Max	BR = C
Tech en arts graphiques	Sal Max	BR = D-
" "	Rém Glob Max	BR = C-
Conduct de véh légers	Sal Min	BR = D
" "	Rém Glob Min	BR = C
Mécanicien d'entretien	Rém Glob Min	BR = C
Electricien d'entretien	Rém Glob Moyenne	BR = E

Bien que le CV et le BR soient des mesures utiles pour déterminer la précision et l'exactitude des données estimées obtenues, il est à noter que ces deux mesures ne sont pas des indicateurs de représentativité. Le CV, par exemple, ne tient pas compte des strates dans lesquelles il n'y a pas d'emploi-repère. Prenons un exemple fictif d'un emploi-repère qui n'est représenté que dans une strate. Si les données de cet emploi-repère étaient très proches, le CV serait très petit, ce qui implique, dans la présentation du CRSMT, qu'elles sont acceptables. Mais il pourrait être difficile d'avoir confiance en des résultats qui ne sont basés que sur quelques observations.

Le fait que le CV, par exemple, ne donne aucune indication sur la fiabilité des emplois-repères quand aucun n'existe dans une strate démontre qu'il ne peut être utilisé comme seul critère d'acceptation en ce qui concerne les emplois repères. Le CV et le BR ne mesurent que les erreurs d'échantillonnage et d'estimation. L'utilisateur doit donc se baser sur d'autres critères pour estimer le degré auquel les emplois repères sont représentés à travers la province, ou leur principaux secteurs ou régions. Le recensement est la meilleure façon d'assurer la couverture d'un certain nombre d'emplois repères. Mais même dans le cas d'un recensement complet et entièrement fiable, l'utilisateur a besoin d'un minimum d'informations pour qu'il puisse interpréter et utiliser les données.

Les critères ou normes d'acceptation dépendent des utilisateurs. Puisque la tâche de l'IRIR est de comparer la rémunération des secteurs publics et parapublics par rapport à ceux des "autres salariés québécois", plus un emploi-repère est représenté dans plusieurs strates, plus fiables et précises seront les données. Ainsi, on peut suggérer que le nombre d'employés et d'établissements devrait atteindre un certain niveau dans toutes les catégories: secteurs d'activité économique, tailles des établissements et régions géographiques. La définition même d'"emploi-repère" serait alors mieux respectée.

Par exemple, si on regroupe les strates de la façon suivante:

1. Secteurs d'activités économiques ou groupement des secteurs:

   Manufacturiers

   Transports

   Finances & Commerce

   Administration locale

   Administration fédérale: emplois seuls

   Services moins Education & Santé

2. Tailles d'établissement (200-499, 500-999, 1000+)

3. Régions géographiques.

Le tableau suivant montre le nombre d'emplois-repères qui se trouvent dans les strates visées ci-haut, selon le nombre d'établissements et d'effectifs de chaque strate. Par exemple, il y a 33 emplois-repères qui ont été observés au moins 10 fois et qui se trouvent dans au moins 3 établissements, dans chacune des strates. Seulement 7 emplois-repères se trouvent dans au moins 15 établissements dans chacune des strates.

Nombre d'emplois-repères retenus selon les critères établissements/strates et effectifs/strates

Etabl. par strate	Observations (effectifs) par strate				
	⩾1	⩾10	⩾20	⩾25	⩾30
⩾3	37	33	21	29	17
⩾5	29	29	21	19	17
⩾10	10	10	10	10	10
⩾15	7	7	7	7	7

Plus le nombre d'établissements et d'observations augmente, moins les emplois-repères sont représentés à travers le marché défini par cette enquête. Ainsi pour avoir cinq (5) établissements par regroupement de strates, et pour avoir en moyenne 5 emplois-repères par établissement (c'est-à-dire 25 observations au total) afin d'avoir une représentation des salaires, seulement 19 (24%) des emplois-repères seraient retenus.(1)

Ceci n'est qu'un exemple pour illustrer la représentativité des emplois-repères. D'autres facteurs peuvent aussi entrer en ligne de compte:

1. Quelques emplois-repères, par exemple celui de l'agent de l'information, se trouvent majoritairement dans le secteur public. Des critères particuliers à ce type d'emploi doivent être formulés.
2. Ce test s'est effectué sur la base des chiffres absolus. Il serait possible aussi de se servir d'une base proportionnelle, par exemple entre le nombre d'emplois existants dans les différents secteurs, tailles, régions, etc.

---

(1) Il est à noter qu'on ne possède pas d'informations sur le nombre d'employés par établissement. Dans le cas du gouvernement fédéral, on ne compte que les employés puisqu'il n'y a qu'un établissement. Les emplois spécifiques aux municipalités n'ont pas été inclus.

Quoique ce test ne donne qu'une indication d'une façon dont la représentativité des emplois-repères peut être évaluée, il suggère que le CV et le BR ne sont pas des critères suffisants pour accepter ou rejeter un emploi-repère parce que celui-ci pourrait avoir des CV et BR acceptables, mais n'être représenté que dans un nombre restreint de strates.

5.3 Avantages/Inconvénients

Avantage

1. Le CV et le BR peuvent servir à indiquer si un emploi-repère doit être rejeté à cause d'un manque de précision ou d'exactitude statistique.

Inconvénients

1. Le lecteur doit choisir ses critères ou normes d'acceptation en ce qui concerne les emplois-repères. En plus, aucune donnée n'est présentée publiquement par le CRSMT sur le nombre d'employés par catégorie d'emploi ventilée en strates. Même avec des données plus détaillées, d'autres informations sur la ventilation des employés par rapport aux établissements sont manquantes.
2. Le CV et le BR sont difficiles à comprendre et par conséquent difficiles à utiliser. Ils devraient être mieux expliqués dans la présentation du CRSMT.

Le CV et le BR sont des indicateurs valables de précision et de l'exactitude statistique des données échantillonnées par méthode aléatoire. Il est à noter que le CRSMT est le premier dans le domaine des enquêtes en rémunération en Amérique du Nord à développer et exploiter le concept du biais relatif et à publier les résultats. Toutefois, le CRSMT devrait présenter plus de données pour aider l'utilisateur à évaluer la représentativité des emplois-repères et inclure aussi le nombre d'employés réels et estimés dans les stratifications ventilées.

CONCLUSION

L'établissement comme unité de base pose quelques problèmes dans les cas des entreprises multi-tailles, mais l'univers est assez grand et fiable pour qu'on puisse se concentrer sur d'autres aspects ayant plus d'impact sur la fiabilité des résultats.

La méthode d'échantillonnage optimale et proportionnelle de Neyman répond bien aux contraintes de ce genre d'enquête statistique. Cependant, le CRSMT n'a pas vérifié la représentativité des emplois-repères des établissements enquêtés par rapport aux emplois-repères des établissements non-enquêtés.

La pondération des données pose également un problème dans cette enquête. Les résultats peuvent être biaisés parce que le CRSMT émet l'hypothèse que les faillites ou les changements de taille à l'intérieur d'un échantillon se trouvent dans les mêmes proportions dans la strate entière. Les établissements recensés sont aussi pondérés lors d'un refus. Dans ces deux cas, le CRSMT suppose que les variances des non-réponses sont les mêmes que celles des taux moyens de salaires, ce qui peut créer des résultats non fiables.

Le CV et le BR sont de bons indicateurs de la précision et de l'exactitude des estimations de données. Il n'y a cependant aucun indicateur du degré de représentativité des emplois-repères à travers les strates.

En somme le CRSMT a adopté une méthodologie qui est dans l'ensemble rigoureuse et acceptable sur le plan statistique. Certaines de ses prémisses restent toutefois à vérifier avant que cette méthode puisse être acceptée en matière de rémunération. Mais même après vérification, il restera à savoir si les parties syndicales et patronales des secteurs public et parapublic du Québec sont prêtes à accepter l'utilisation d'une source de comparaison dont les marges d'erreurs implicites peuvent être de nature à rendre la comparaison et le calcul de l'écart public/autre non concluante..

** Le rapport de l'IRIR "Analyse de l'enquête du CRSMT" soulève d'autres aspects de cette enquête, notamment en ce qui concerne les appariements d'emploi.

REFERENCES

William C. COCHRAN, Sampling techniques, 3rd Ed., John Wiley and Sons, New York, 1977, 428 p.

SHEAFFER, MENDENHALL & OTT. Elementary Survey Sampling 2nd Ed., Duxbury Press, Belmont CA, 1979, 278 p.

ANNEXE 1

La prochaine page explique les concepts du niveau de confiance, du coefficient de variation et du biais relatif dans l'exemple de la page 3.

Source:
Emploi, gains et durée du travail
Statistique Canada
Catalogue 72-002
décembre, 1985
185-186 pp.

La deuxième page offre une représentation visuelle du coefficient de variation (CV) et du biais relatif (BR) fourni par le CRSMT.

Fiabilité des estimations

En ce qui concerne les estimations basées sur un sondage, il existe deux types d'erreurs possibles, soit les erreurs d'échantillonnage et les erreurs non liées à l'échantillonnage.

Les erreurs d'échantillonnage proviennent de ce que les observations se font à partir d'un échantillon au lieu de l'univers complet. Les estimations basées sur un échantillon diffèrent des statistiques que l'on obtiendrait d'un recensement complet effectué au moyen des mêmes questionnaires, des mêmes directives, des mêmes interviewers et des mêmes techniques de traitement. Cette différence, lorsqu'il s'agit d'une estimation particulière, s'appelle erreur d'échantillonnage de l'estimation.

Les erreurs non dues à l'échantillonnage ne se rattachent pas à l'échantillon et peuvent se produire tout au long d'une enquête pour plusieurs raisons. Les cas de non-réponse, les anomalies de couverture et de classification, les différences d'interprétation des questions, la communication de renseignements inexacts par les répondants et les fautes d'entrée, de codage et de traitement des données sont autant d'erreurs non attribuables à l'échantillonnage. Il faut noter que des erreurs de ce genre peuvent aussi se produire dans les recensements. Pour l'enquête sur l'emploi, la rémunération et les heures de travail, nous avons tenté d'en réduire le risque par une formulation soignée des questionnaires, par une vérification rigoureuse des chiffres déclarés, par le travail de suivi, par l'imputation de données aux unités non répondantes et par un contrôle attentif des techniques de traitement.

Indicateurs des erreurs d'échantillonnage

L'erreur d'échantillonnage que nous venons de décrire reste inconnue. On peut cependant l'estimer à partir de l'échantillon, en utilisant une mesure appelée **écart-type** d'une estimation de sondage. Lorsque l'écart-type est exprimé en pourcentage de l'estimation, on l'appelle écart-type relatif ou coefficient de variation.

Dans la présente publication, nous utilisons des symboles littéraux de coefficient de variation (voir tableaux A à F à la fin de cette note explicative), pour indiquer le degré de fiabilité des estimations; cette symbolisation repose sur la moyenne des coefficients de variation de 12 mois. Les intervalles de confiance des coefficients de variation et les symboles correspondants sont précisés dans le tableau suivant.

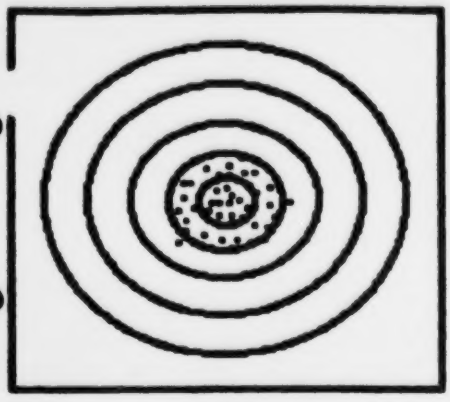
INDICATEURS LITTÉRAUX DU COEFFICIENT DE VARIATION

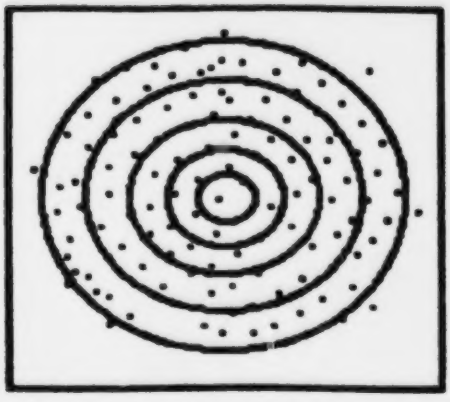
Indicateur littéral	Intervalle de confiance du coefficient de variation en pourcentage
A	0 - 1.0
B	1.1 - 2.5
C	2.6 - 5.0
D	5.1 - 10.0
E	10.1 - 15.0
F	15.1 - 25.0
G	> 25.0

L'estimation de l'échantillon et son écart-type (établi d'après le coefficient de variation) peuvent servir à construire un intervalle à l'intérieur duquel s'inscrira, sous réserve d'une limite de confiance prédéterminée, la valeur inconnue que l'on obtiendrait par recensement. L'échantillon sur lequel s'appuie notre enquête n'est qu'un des multiples fragments de même taille que nous aurions pu choisir en utilisant le même plan. Les estimations tirées d'autres échantillons du même univers différeraient les unes des autres. Si l'on construisait des intervalles allant de deux écarts-types sous l'estimation à deux écarts-types au-dessus de l'estimation, en utilisant chacune des estimations possibles (selon l'échantillon choisi), environ 19 intervalles sur 20 contiendraient alors la valeur que donnerait un recensement. Ce genre d'intervalle se nomme intervalle de confiance à 95%. Prenons par exemple comme estimation le nombre total de salariés de la construction pour l'ensemble du Canada, et supposons que ce nombre soit 450,000. Si l'indicateur de fiabilité de cette estimation est A, et si son coefficient de variation s'établit à 1%, alors l'écart-type de l'estimation sera de 4,500 salariés. Dans ce cas, on peut donc dire avec 95% de confiance que le nombre total de salariés de la construction, obtenu par recensement, se situe dans l'intervalle 441,000 à 459,000 salariés.

En général, on s'attend à ce que les estimations des variations mensuelles soient plus fiables que les estimations de niveaux, par suite de l'ample chevauchement des unités déclarantes au sein des échantillons mensuels, et de la forte corrélation qui existe entre les valeurs déclarées pour deux mois consécutifs.

ANNEXE I

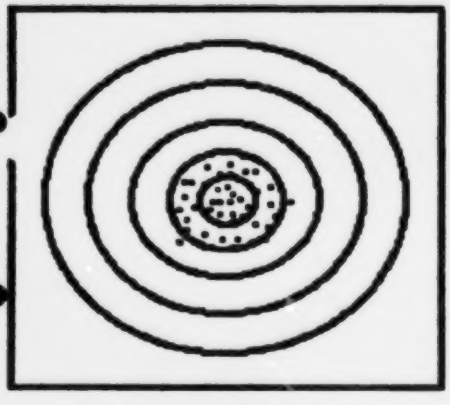


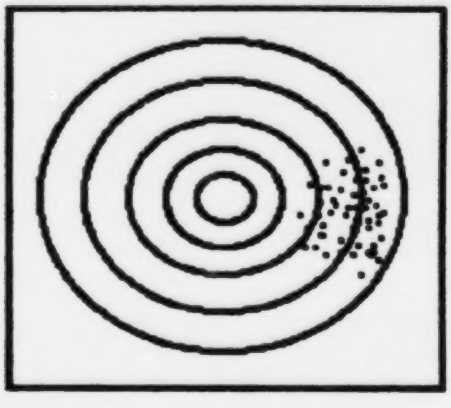


\+ A

— E

Précision (C.V.)





\+ A

— D

Exactitude (B.R.)

Notes:
- Les deux mesures qualifient les données à leur manière sur un aspect particulier.
- Toutes les combinaisons sont possibles.

SOURCE: CRSMT